

Non-contiguous finished genome sequence and description of *Gorillibacterium massiliense* gen. nov, sp. nov., a new member of the family *Paenibacillaceae*

Mamadou Bhoye Keita¹, Roshan Padhmanabhan¹, Aurélia Caputo¹, Catherine Robert¹, Eric Delaporte², Didier Raoult^{1,3}, Pierre-Edouard Fournier¹, and Fadi Bittar^{1*}

¹URMITE, Aix-Marseille Université, Faculté de médecine, Marseille, France

²IRD, University Montpellier 1, Montpellier, France

³King Fahad Medical Research Center, King Abdul Aziz University, Jeddah, Saudi Arabia

*Correspondence: Fadi Bittar (fadi.bittar@univ-amu.fr)

Keywords: *Gorillibacterium massiliense*, genome, culturomics, taxono-genomics

Strain G5^T gen. nov., sp. nov. is the type strain of *Gorillibacterium massiliense*, a newly proposed genus within the family *Paenibacillaceae*. This strain, whose genome is described here, was isolated in France from a stool sample of a wild *Gorilla gorilla* subsp. *gorilla* from Cameroon. *G. massiliense* is a facultatively anaerobic, Gram negative rod. Here we describe the features of this bacterium, together with the complete genome sequence and annotation. The 5,546,433 bp long genome (1 chromosome but no plasmid) contains 5,145 protein-coding and 76 RNA genes, including 69 tRNA genes.

Introduction

Strain G5^T (= CSUR P290 = DSM 27179) is the type strain of *Gorillibacterium massiliense* gen. nov., sp. nov. This bacterium which is proposed to belong to the family *Paenibacillaceae*, is a Gram-negative, flagellated, facultative anaerobic, indole-negative bacillus that was isolated from a fecal sample of a wild western lowland gorilla from Cameroon, through a culturomics study of the bacterial diversity of the feces of wild gorillas. This technique was used successfully to explore the human gut microbiota allowing the isolation of many new species and genera [1-3].

The newly proposed strategy of applying high throughput genome sequencing, MALDI-TOF spectral analysis of cellular proteins, coupled with more traditional methods of phenotypic characterization has been demonstrated as a useful approach for the description of new bacterial taxa [4-15]. A principle advantage is that this method circumvents the vagaries of methods that rely mainly on DNA-DNA hybridization to delineate species. Here, we applied this polyphasic approach to describe *G. massiliense* gen. nov., sp. nov. strain G5^T.

The family *Paenibacillaceae* [16] belongs to the phylum *Firmicutes* and includes the 9 following genera [17]: *Paenibacillus* [18,19], *Ammoniphilus* [20], *Aneurinibacillus* [21], *Brevibacillus* [21], *Thermobacillus* [22], *Fontibacillus* [23], *Cohnella* [24], *Saccharibacillus* [25] and *Oxalophagus* [26].

Members belonging to this family were isolated mainly from soil, roots, blood, feces and other sources [16]. To the best of our knowledge, this is the first report of the isolation of a novel genus from the fecal flora of a gorilla.

Here we present a summary classification and a set of features for *G. massiliense* gen. nov., sp. nov. strain G5^T (= CSUR P290 = DSM 27179) together with the description of the complete genomic sequencing and its annotation. These characteristics support the circumscription of a novel genus, *Gorillibacterium* gen. nov. within the family *Paenibacillaceae*, with *Gorillibacterium massiliense* gen. nov., sp. nov. as the type species.

Classification and features

In July 2011, a fecal sample was collected from a wild *Gorilla gorilla* subsp. *gorilla* near Minton, a village in the south-central part of the DJA FAUNAL Park (Cameroon). The collection of the stool sample was approved by the Ministry of Scientific Research and Innovation of Cameroon. No experiments were conducted on this gorilla. The fecal specimen was preserved at -80°C after collection and sent to Marseille. Strain G5^T (Table 1) was isolated in August 2012 by aerobic cultivation at 37°C on sterilized soil medium (12 g of soil (Latitude: N 43° 17' 20.151"; Longitude: E 5° 24' 15.3822") / agar (14g/l). This strain exhibited a 93.72% 16S rRNA nucleotide sequence

Gorillibacterium massiliense

similarity with *Paenibacillus turicensis*, the phylogenetically closest validly published *Paenibacillus* species (Figure 1). This value was lower than the 95.0% 16S rRNA gene sequence

threshold recommended by Stackebrandt and Ebers to delineate a new genus without carrying out DNA-DNA hybridization [37].

Table 1. Classification and general features of *Gorillibacterium massiliense* strain G5^T

| MIGS ID | Property | Term | Evidence code ^a |
|----------|------------------------|---|----------------------------|
| | Current classification | Domain Bacteria | TAS [27] |
| | | Phylum <i>Firmicutes</i> | TAS [28-30] |
| | | Class <i>Bacilli</i> | TAS [31,32] |
| | | Order <i>Bacillales</i> | TAS [33,34] |
| | | Family <i>Paenibacillaceae</i> | TAS [16,32] |
| | | Genus <i>Gorillibacterium</i> | IDA |
| | | Species <i>Gorillibacterium massiliense</i> | IDA |
| | | Type strain G5T | IDA |
| | Gram stain | Negative | IDA |
| | Cell shape | rod | IDA |
| | Motility | non-motile | IDA |
| | Sporulation | non-sporulating | IDA |
| | Temperature range | mesophilic | IDA |
| | Optimum temperature | 37°C | IDA |
| MIGS-6.3 | Salinity | no Growth in BHI medium + 5% NaCl | IDA |
| MIGS-22 | Oxygen requirement | facultative anaerobic | IDA |
| | Carbon source | varied (see Table 2) | IDA |
| | Energy source | Chemoorganoheterotrophic | IDA |
| MIGS-6 | Habitat | gorilla gut | IDA |
| MIGS-15 | Biotic relationship | free living | IDA |
| | Pathogenicity | Unknown | NAS |
| | Biosafety level | 2 | NAS |
| MIGS-14 | Isolation | Gorilla feces | IDA |
| MIGS-4 | Geographic location | Cameroon | IDA |
| MIGS-5 | Sample collection time | July 2011 | IDA |
| MIGS-4.1 | Latitude | 2.783938 | IDA |
| MIGS-4.1 | Longitude | 13.030472 | IDA |
| MIGS-4.3 | Depth | surface | IDA |
| MIGS-4.4 | Altitude | > 600 m above sea level | IDA |

^aEvidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [35]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.

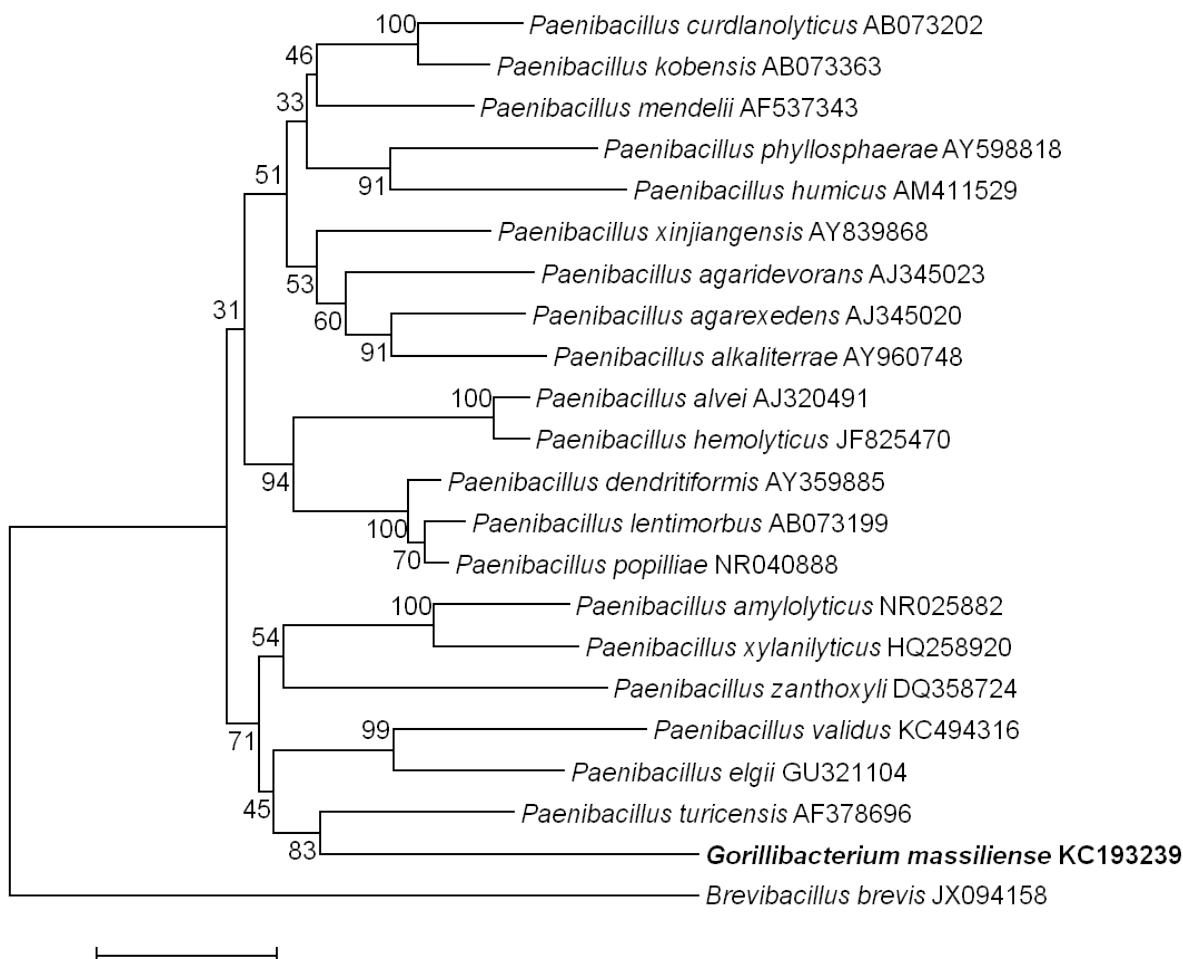


Figure 1. Phylogenetic tree highlighting the position of *Gorillibacterium massiliense* strain G5^T relative to other type strains within the *Paenibacillaceae* family. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTAL X (V2), and phylogenetic inferences obtained using the maximum-likelihood method within the MEGA 5 software [36]. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Brevibacillus brevis* was used as out-group. The scale bar represents a 2% nucleotide sequence divergence.

Different growth temperatures (25, 30, 37, 45°C) were tested. No growth occurred at 45°C, growth occurred between 25° and 37°C, and optimal growth was observed at 37°C. Colonies were bright grey with a diameter of 1.0 mm on 5% blood-enriched Columbia agar. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and under aerobic conditions, with or without 5% CO₂. Growth was observed under anaerobic and microaerophilic conditions, but optimal growth was obtained aerobically. Moreover, the Gram staining showed Gram-negative rod (Figure 2). A motility test produced a negative result. Cells grown on agar did not sporulate and the rods exhibited peritrichous flagella and had a mean length of 1.75 µm and a mean diameter of 0.67 µm as determined by negative staining transmission electron microscopy (Figure 3).

Strain G5^T exhibited catalase activity but not oxidase activity. Using the API 50CH system (BioMérieux), a positive reaction was obtained for D-xylose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, aesculin, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, inulin, D-melezitose, D-raffinose, glycogen, gentiobiose, D-turanose, Methyl-α-D-glucopyranoside and hydrolysis of starch. A weak positive reaction was observed for L-arabinose. A negative reaction was observed for glycerol, ribose, D-galactose, L-rhamnose, L-sorbose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl-αD-mannopyranoside, D-arabinose, amygdalin, arbutin, potassium gluconate, potassium 2-cetogluconate, potassium 5-cetogluconate, adonitol and D-tagatose. Using the API ZYM system, positive reactions were obtained only for naphthol-AS-BI-phosphohydrol-ase, α-galactosidase, β-galactosidase, β-glucosidase, arginine

arylamidase and arginine dihydrolase. The production of α -glucosidase, β -glucuronidase, esterase lipase, leucine arylamidase, cystine arylamidase, valine arylamidase, glycine arylamidase, phenylalanine arylamidase, lipase, alkaline phosphatase, acid phosphatase, N-acetyl- β -glucosaminidase and α -chymotrypsin were negative. Urease reaction and reduction of nitrates to nitrogen were also positive. Indole production was negative. *G. massiliense* was sus-

ceptible to ticarcillin, amoxicillin, tobramycin, imipenem, vancomycin and rifampin but resistant to ceftazidime (Caz 30), colistin (CT50) and metronidazole.

When compared with representative species from the family *Paenibacillaceae* [38-42], *G. massiliense* gen. nov., sp. nov. strain G5^T exhibited the phenotypic differences detailed in Table 2.

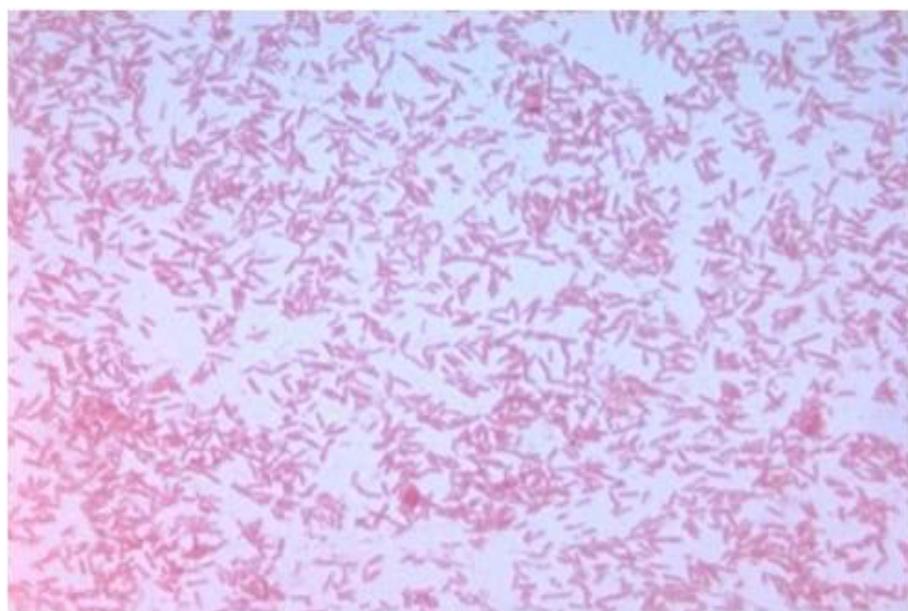


Figure 2. Gram staining of *G. massiliensis* strain G5^T.

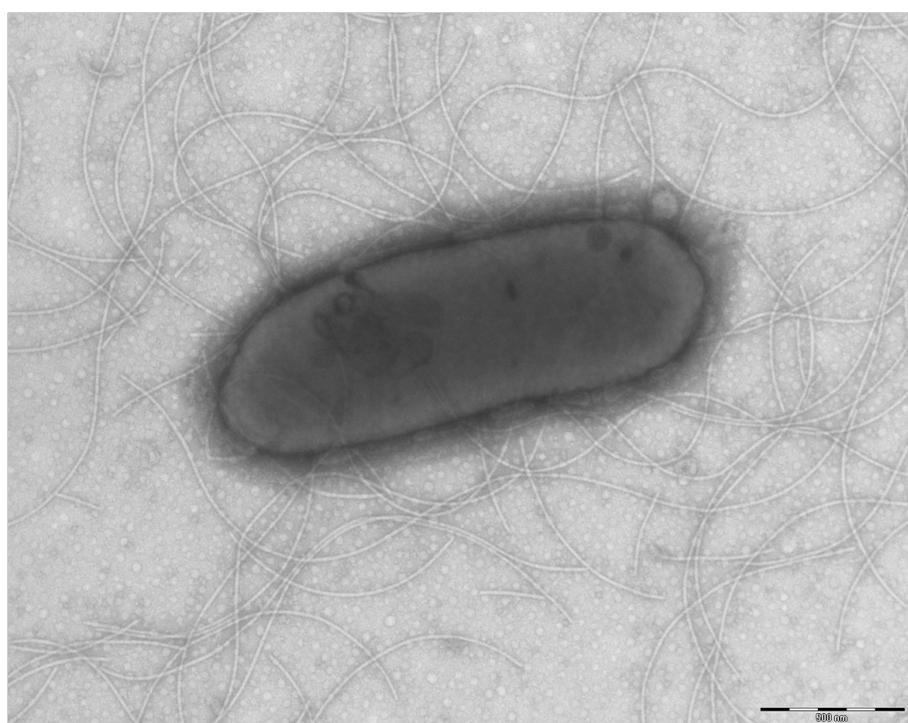


Figure 3. Transmission electron microscopy of *G. massiliense* strain G5^T using a Morgani 268D (Philips) at an operating voltage of 60kV. The scale bar represents 500 nm.

Table 2. Differential phenotypic characteristics between *Gorillibacterium massiliense* gen. nov., sp. nov., strain G5^T and phylogenetically close *Paenibacillaceae* species.

| Characteristic | 1 | 2 | 3 | 4 | 5 |
|----------------------------|-------------|---|------------------------------------|-----------------|-------------------|
| Gram stain | - | + | var | var | +/var |
| Motility | - | + | + | + | + |
| Endospore formation | - | + | + | + | + |
| | | Human: valve of cerebrospinal fluid shunt | Roots of <i>Perilla frutescens</i> | Honeybee larvae | Environment: soil |
| Isolated from | Gorilla gut | | | | |
| Production of | | | | | |
| Catalase | + | - | + | + | + |
| Oxidase | - | - | - | + | + |
| Nitrate reductase | + | - | + | - | + |
| Urease | + | - | + | na | - |
| Indole | - | - | + | + | - |
| Utilization of: | | | | | |
| Glycerol | - | - | var | - | - |
| D-xylose | + | + | var | - | - |
| D-glucose | + | + | + | + | - |
| D-fructose | + | + | w | na | - |
| D-mannose | + | + | + | - | - |
| Methyl- αD-mannopyranoside | - | - | na | na | - |
| N-acetylglucosamine | + | + | + | na | + |
| Aesculin | + | + | + | na | + |
| Salicin | + | + | - | - | - |
| D-celllobiose | + | + | + | - | + |
| D-maltose | + | + | + | na | - |
| D-lactose | + | + | + | na | - |
| D-melibiose | + | + | na | + | - |
| D-saccharose | + | + | na | - | - |
| D-trehalose | + | - | + | - | - |
| D-melezitose | + | - | na | - | - |
| D-raffinose | + | + | na | + | - |
| Starch | + | + | + | + | - |
| Glycogen | + | + | w | - | - |
| β-Gentibiose | + | + | w | na | - |
| L-arabinose | w | + | - | - | - |
| Ribose | - | + | + | na | - |
| D-galactose | - | + | + | na | + |
| D-mannitol | - | - | + | - | + |
| Potassium gluconate | - | - | + | - | w |
| Amygdalin | - | + | - | - | - |

Strain: 1, *Gorillibacterium massiliense* G5^T; 2, *Paenibacillus turicensis* MOL722^T; 3, *Paenibacillus elgii* SD17^T; 4, *Paenibacillus alvei* BCRC 11220^T; 5, *Brevibacillus brevis* NBRC 15304^T.

-: negative result, +: positive result, var: variable, na: data not available, w: weak positive result

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [15] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). Twelve distinct deposits were done for strain G5^T from 12 isolated colonies. The 12 G5^T spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against 6,252 bacterial spectra used as reference data, in the BioTyper database. A score enabled the pre-

sumptive identification of the isolated based on the following heuristic species: a score ≥ 2 with a validated species enabled the identification at the species level, a score ≥ 1.7 but < 2 enabled the identification at the genus level; and a score < 1.7 did not enable any identification. For strain G5^T, a significant score was not obtained, suggesting it was not a member of any known species or genus. We incremented our database with the spectrum from strain G5^T (Figure 4). Spectrum differences with other of *Paenibacillaceae* family are shown in Figure 5.

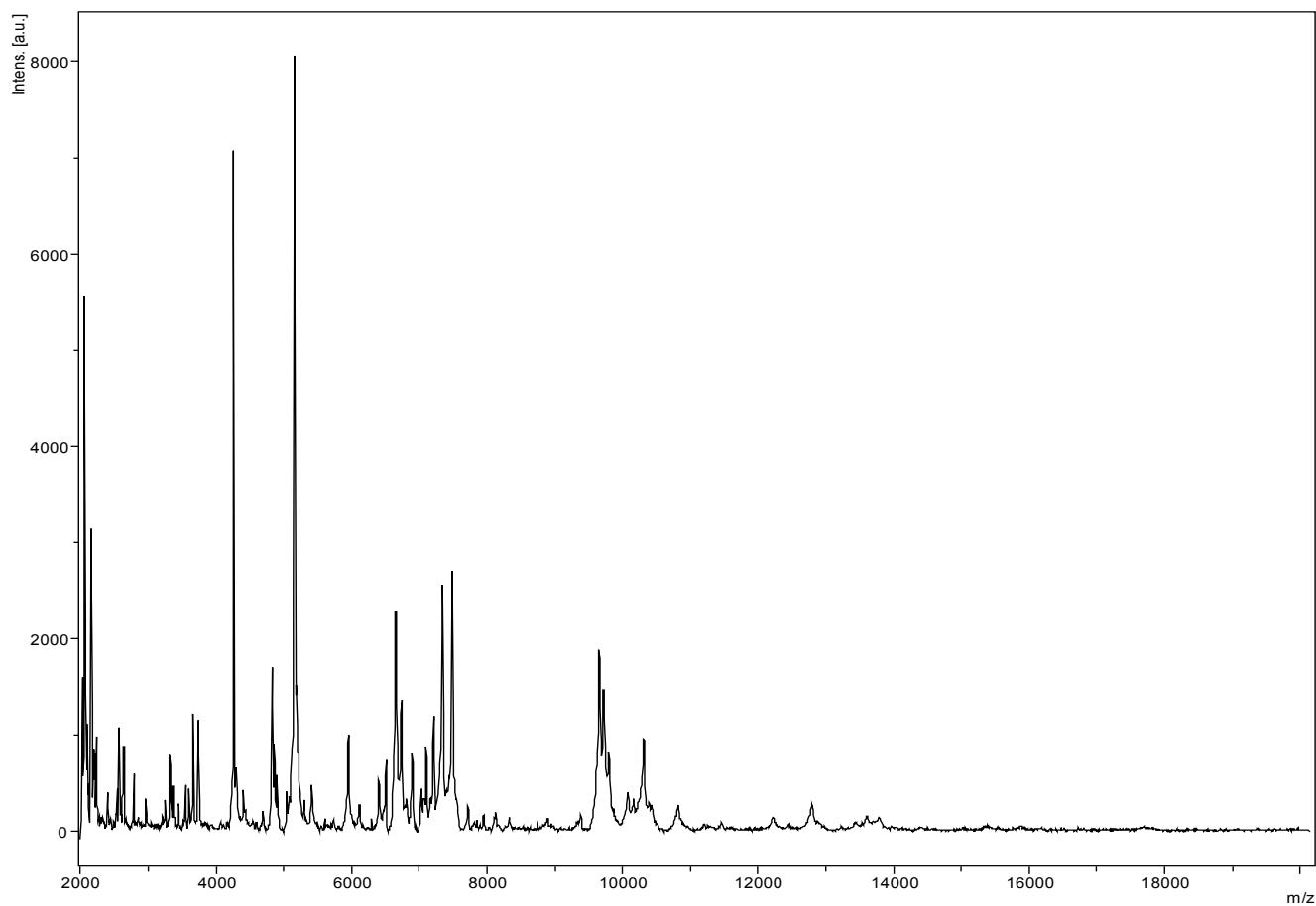


Figure 4. Reference mass spectrum from *G. massiliense* strain G5^T. Spectra from 16 individual colonies were compared and a reference spectrum was generated.

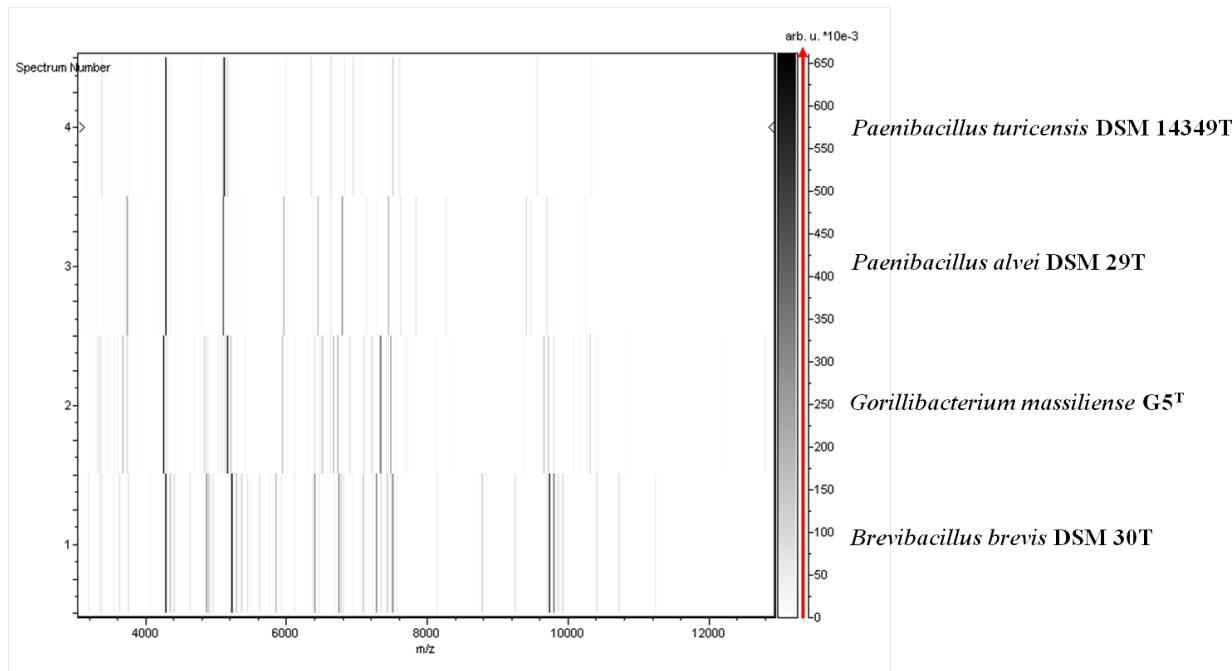


Figure 5. Gel view comparing *Gorillibacterium massiliensis* gen. nov., sp. nov. strain G5^T spectra with other members of the *Paenibacillaceae* family. The Gel View displays the raw spectra of all loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Gray scale scheme code. The color bar and the right y-axis indicate the relation between the color a peak is displayed with and the peak intensity in arbitrary units. Displayed species are indicated on the left.

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its phylogenetic position and 16S rRNA similarity to other members of the family *Paenibacillaceae*, and is part of a “culturomics” study of the gorilla flora aiming at isolating all bacterial species within gorilla feces. It was the 81st genome of the *Paenibacillaceae* family and

the first genome of *Gorillibacterium massiliense* gen. nov., sp. nov. A summary of the project information is shown in Table 3. The Genbank accession number is CBQR000000000 and consists of 176 large contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [43].

Table 3. Project information

| MIGS ID | Property | Term |
|-----------|----------------------|-------------------------------------|
| MIGS-31 | Finishing quality | High-quality draft |
| MIGS-28 | Libraries used | 454 paired-end 3- kb libraries |
| MIGS-29 | Sequencing platform | 454 GS FLX Titanium |
| MIGS-31.2 | Sequencing coverage | 25.71× |
| MIGS-30 | Assemblers | Newbler version 2.5.3 |
| MIGS-32 | Gene calling method | Prodigal |
| | EMBL Date of Release | August 07, 2013 |
| | EMBL ID | CBQR000000000 |
| MIGS-13 | Project relevance | Study of the gorilla gut microbiome |

Growth conditions and DNA isolation

Gorillibacterium massiliense gen. nov., sp. nov., strain G5^T (= CSUR P290 = DSM 27179) was grown aerobically on 5% sheep blood-enriched Columbia agar at 37°C. Four petri dishes were spread and resuspended in 3×500µL of TE buffer and stored at 80°C. Then, 500 µL of this suspension were thawed, centrifuged 3 minutes at 10,000 rpm and resuspended in 3×100 µL of G2 buffer (EZ1 DNA Tissue kit, Qiagen). A first mechanical lysis was performed by glass powder on the Fastprep-24 device (Sample Preparation system, MP Biomedicals, USA) using 2×20 seconds cycles. DNA was then treated with 2.5µg/µL lysozyme (30 minutes at 37°C) and extracted using the BioRobot EZ1 Advanced XL (Qiagen). The DNA was then concentrated and purified using the Qiaamp kit (Qiagen). The yield and the concentration were measured by the Quant-it Picogreen kit (Invitrogen) on the Genios Tecan fluorometer at 50ng/µL.

Genome sequencing and assembly

The paired-end library was prepared with 5 µg of bacterial DNA using DNA fragmentation on a Covaris S-Series (S2) instrument (Woburn, Massachusetts, USA) with an enrichment size at 4.5kb. DNA fragmentation was visualized with an Agilent 2100 BioAnalyzer on a DNA labchip 7500. The library was constructed according to the 454 GS FLX Titanium paired-end protocol (Roche). Circularization and nebulization were performed and generated a pattern with an op-

timum at 510 bp. After PCR amplification through 17 cycles followed by double size selection, the single stranded paired-end library was quantified using a BioAnalyzer 2100 on a RNA pico 6000 labchip at 68 pg/µL. The library concentration equivalence was calculated as 2.45E+08 molecules/µL. The library was stored at -20°C until further use.

The paired-end library was clonally amplified with 0.25 cpb and 0.5 cpb in 2 emPCR reactions with the GS Titanium SV emPCR Kit (Lib-L) v2 (Roche). The yield of the emPCR was respectively of 5 and 6% as expected of the yield ranging from 5 to 20% recommended by the Roche procedure.

Approximately 790,000 beads were loaded twice (i.e. two runs were performed using the same paired-end library) on a ¼ region of the GS Titanium PicoTiterPlate PTP Kit 70×75 and sequenced with the GS FLX Titanium Sequencing Kit XLR70 (Roche). The two runs were performed overnight and then analyzed on the cluster through the gsRunBrowser and Newbler assembler (Roche). A total of 387,157 passed filter wells were obtained and generated 142.7 Mb of sequences with a length average of 369 bp. The passed filter sequences were assembled using Newbler with 90% identity and 40-bp as overlap. The final assembly identified 12 scaffolds with 176 large contigs (>1.5kb), generating a genome size of 5.5 Mb which corresponds to a genome coverage of 25.71×.

Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal [44] with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [45] and the Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAscanSE tool [46] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer [47] and BLASTn against the GenBank database. ORFans were identified if their BLASTP *E*-value was lower than 1e-03 for alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an *E*-value of 1e-05.

To estimate the mean level of nucleotide sequence similarity at the genome level between *G. massiliense* and another 2 members of the family *Paenibacillaceae* and *Brevibacillus brevis*, we use the Average Genomic Identity of Orthologous gene Sequences (AGIOS), a custom application we developed. Briefly, the AGIOS software com-

bines the Proteinortho software [48] for detecting orthologous proteins between genomes compared two by two, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm.

Genome properties

The genome is 5,546,433 bp long with a 50.39% G+C content (Figure 6 and Table 4). It is composed of 189 Contigs (176 large contigs, 12 scaffolds). Of the 5,221 predicted genes, 5,145 were protein-coding genes, and 76 were RNAs (1 gene is 16S rRNA, 1 gene is 23S rRNA, 5 genes are 5S rRNA, and 69 are tRNA genes). A total of 3,865 genes (75.12%) were assigned a putative function (by cogs or by NR blast). In addition, 272 genes were identified as ORFans (5.29%). The remaining genes were annotated as hypothetical proteins (680 genes => 13.22%). The distribution of genes into COGs functional categories is presented in Table 5. The properties and the statistics of the genome are summarized in Table 4 and 5.

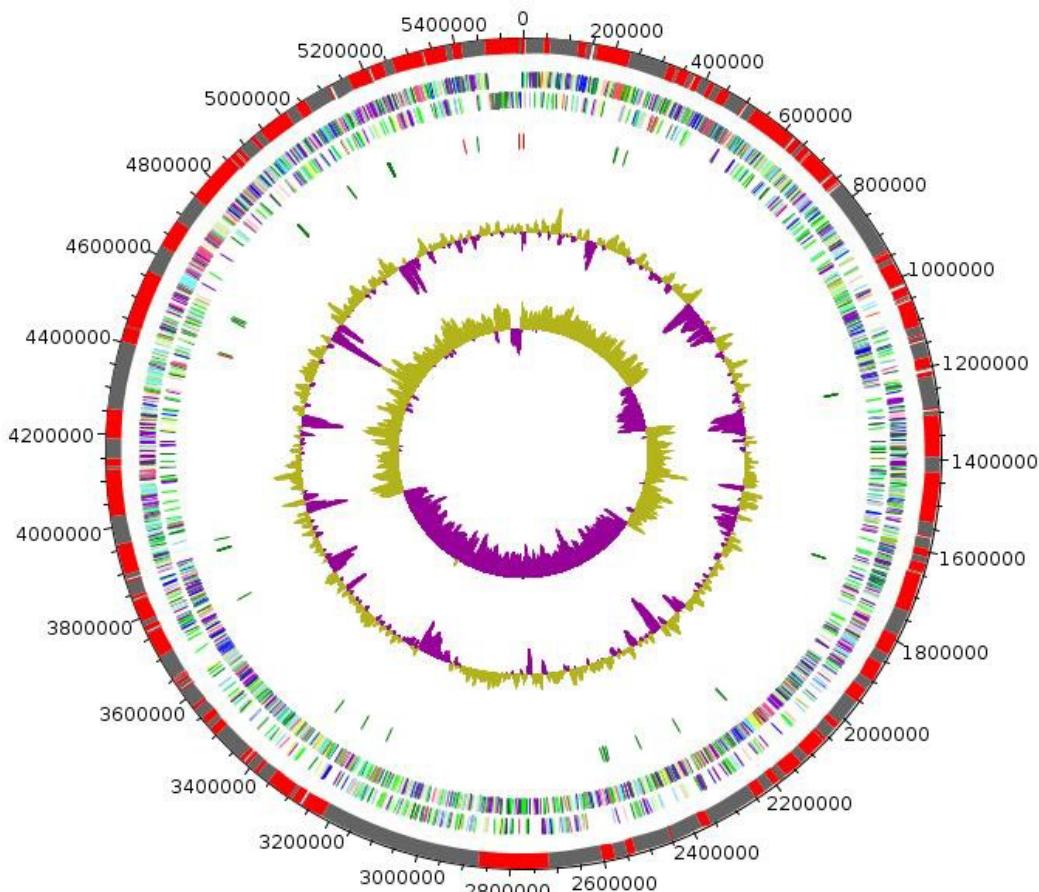


Figure 6. Graphical circular map of the chromosome. From outside to the center: Genes on the forward strand colored by COG categories (only genes assigned to COG), genes on the reverse strand colored by COG categories (only gene assigned to COG), RNA genes (tRNAs green, rRNAs red), G+C content and GC skew. Purple and olive indicating negative and positive values, respectively.

Table 4. Nucleotide content and gene count levels of the chromosome

| Attribute | Value | % of total ^a |
|----------------------------------|-----------|-------------------------|
| Genome size (bp) | 5,546,433 | 100 |
| DNA G+C content (bp) | 2,794,611 | 50.39 |
| DNA coding region (bp) | 4,888,209 | 88.13 |
| Total genes | 5,221 | 100 |
| RNA genes | 76 | 1.46 |
| Protein-coding genes | 5,145 | 98.54 |
| Genes with function prediction | 3,865 | 75.12 |
| Genes assigned to COGs | 3,881 | 75.43 |
| Genes with peptide signals | 709 | 13.78 |
| Genes with transmembrane helices | 1,267 | 24.63 |

^a The total is based on either the size of the genome in base pairs or the total number of protein-coding genes in the annotated genome

Table 5. Number of genes associated with the 25 general COG functional categories

| Code | Value | % age ^a | Description |
|------|-------|--------------------|--|
| J | 201 | 3.91 | Translation, ribosomal structure and biogenesis |
| A | 0 | 0 | RNA processing and modification |
| K | 483 | 9.39 | Transcription |
| L | 166 | 3.23 | Replication, recombination and repair |
| B | 1 | 0.02 | Chromatin structure and dynamics |
| D | 43 | 0.84 | Cell cycle control, mitosis and meiosis |
| Y | 0 | 0 | Nuclear structure |
| V | 119 | 2.31 | Defense mechanisms |
| T | 313 | 6.08 | Signal transduction mechanisms |
| M | 205 | 3.98 | Cell wall/membrane biogenesis |
| N | 73 | 1.42 | Cell motility |
| Z | 5 | 0.1 | Cytoskeleton |
| W | 0 | 0 | Extracellular structures |
| U | 50 | 0.97 | Intracellular trafficking and secretion |
| O | 121 | 2.35 | Posttranslational modification, protein turnover, chaperones |
| C | 180 | 3.5 | Energy production and conversion |
| G | 560 | 10.88 | Carbohydrate transport and metabolism |
| E | 355 | 6.9 | Amino acid transport and metabolism |
| F | 93 | 1.81 | Nucleotide transport and metabolism |
| H | 130 | 2.53 | Coenzyme transport and metabolism |
| I | 108 | 2.1 | Lipid transport and metabolism |
| P | 248 | 4.82 | Inorganic ion transport and metabolism |
| Q | 113 | 2.2 | Secondary metabolites biosynthesis, transport and catabolism |
| R | 648 | 12.59 | General function prediction only |
| S | 328 | 6.38 | Function unknown |
| - | 1,264 | 24.57 | Not in COGs |

^a The total is based on the total number of protein-coding genes in the annotated genome

Genomic comparison of *G. massiliense* and other members of the family Paenibacillaceae

The genome of *G. massiliense* strain G5^T was compared to those of *P. elgii* strain B69, *P. alvei* strain DSM 29 and *B. brevis* strain NBRC 100599 (Table 6A and Table 6B). The draft genome of *G. massiliense* is smaller in size than those of *P. elgii*, *P. alvei* and *B. brevis* (5.54 vs 7.96, 6.83 and 6.3 Mb respectively). *G. massiliense* has a lower G+C content than *P. elgii* (50.39% vs 52.6%) but higher than those of *P. alvei* and *B. brevis* (50.39% vs 45.9% and 47.3% respectively). The protein content of *G. massiliense* is lower than those of *P. elgii*, *P. alvei* and *B. brevis* (5,146 vs 7,597, 6,823 and 5,946 respectively) (Table 6 and Table 6B). In addition, *G. massiliense* shares 2,122, 1,846 and 1,716 orthologous genes with *P. elgii*, *P. alvei* and *B. brevis*, respectively (Table

6). The nucleotide sequence identity of orthologous genes ranges from 66 to 67.6% among previously published genomes, and from 65.3 to 68.7% between *G. massiliense* and other studied genomes (Table 6A and Table 6B). Table 6 summarizes the number of orthologous genes and the average percentage of nucleotide sequence identity between the different genomes studied.

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Gorillibacterium massiliense* gen. nov., sp. nov., that contains the strain G5^T. This bacterium has been found in stool sample of wild gorilla collected in Cameroon.

Table 6A. Genomic comparison of *G. massiliense* gen. nov., sp. nov., strain G5^T with four other members of the family Paenibacillaceae[†]

| Species | Strain | Genome accession number | Genome size (Mb) | G+C content |
|-------------------------------------|-----------------|-------------------------|------------------|-------------|
| <i>Gorillibacterium massiliense</i> | G5 ^T | CBQR000000000 | 5.54 | 50.39 |
| <i>Paenibacillus elgii</i> | B69 | AFHW000000000 | 7.96 | 52.6 |
| <i>Paenibacillus alvei</i> | DSM 29 | AMBZ000000000 | 6.83 | 45.9 |
| <i>Brevibacillus brevis</i> | NBRC 100599 | AP008955 | 6.3 | 47.3 |

[†]Species and strain names, GenBank genome accession numbers, sizes and G+C contents

Table 6B. Genomic comparison of *G. massiliense* gen. nov., sp. nov., strain G5^T with four other members of the family Paenibacillaceae[†]

| | <i>G. massiliense</i> | <i>P. elgii</i> | <i>P. alvei</i> | <i>B. brevis</i> |
|-----------------------|-----------------------|-----------------|-----------------|------------------|
| <i>G. massiliense</i> | 5,146 | 68.7 | 66.7 | 65.3 |
| <i>P. elgii</i> | 2,122 | 7,597 | 67.6 | 66.4 |
| <i>P. alvei</i> | 1,846 | 2,336 | 6,823 | 66 |
| <i>B. brevis</i> | 1,716 | 2,278 | 1,936 | 5,946 |

[†]Numbers of orthologous protein shared between genomes (lower left triangle), average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (upper right triangle). Bold numbers indicate numbers of proteins per genome.

Description of *Gorillibacterium* gen. nov.

Gorillibacterium (go.ri.li.bac.te.ri'um. gor.il.i NL gen fem, the genus name of the great ape; bac.ter'i.um N.L. neut. n., bacterium a rod; gorillibacterium a rod-shaped bacterium isolated from a gorilla).

Gram-negative rod. Facultatively anaerobic. Mesophilic. Non-motile. Oxidase negative, catalase positive. Positive for urease, nitrate reduction, α - and β -galactosidase, arginine dihydrolase, arginine arylamidase, and β -glucosidase. Habitat: gorilla gut. Type species: *Gorillibacterium massiliense*.

Description of *Gorillibacterium massiliense* gen. nov., sp. nov.

Gorillibacterium massiliense (ma.si.li.en'se. L. gen. neut. n. *massiliense*, of *Massilia*, the ancient Roman name for Marseille, France, where the type strain was isolated).

G. massiliense is Gram-negative rod. Facultatively anaerobic. Mesophilic. Optimal growth is achieved at 37°C. Non-sporulating and non-motile bacterium. Colonies are bright gray and 0.5-1 mm in diameter on blood-enriched Columbia agar. Cells are rod-shaped and have a mean diameter of 0.67 μm and a mean length of 1.75 μm .

Catalase positive, oxidase negative. Using the API 20NE system, positive reactions are observed for nitrate reduction and urease reaction, but indole production was negative. Using the API 50CH system (BioMerieux), a positive reaction was obtained for the fermentation of D-xylose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, aesculin, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, inulin, D-melezitose, D-raffin-

ose, glycogen, gentiobiose, D-turanose, Methyl- α D-glucopyranoside and starch. Negative reactions are observed for glycerol, ribose, D-galactose, L-rhamnose, L-sorbose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl- α D-manno-pyranoside, D-arabinose, amygdalin, arabin, potassium gluconate, potassium 2-cetogluconate, potassium 5-cetogluconate, adonitol and D-tagatose. Using the API ZYM system, positive reactions were observed for the production of naphthol-AS-BI-phosphohydrolase, α -galactosidase, β - galactosidase, β - glucosidase, Arginine arylamidase and Arginine dihydrolase. The production of α - glucosidase, β -glucuronidase, esterase lipase, leucine arylamidase and cystine arylamidase, valine arylamidase, glycine arylamidase, phenylalanine arylamidase, lipase, alkaline phosphatase, acid phosphatase, N-acetyl- β -glucosaminidase and a-chymotrypsin are negative. Susceptible to ticarcillin, amoxicillin, tobramycin, imipenem, vancomycin and rifampin but resistant to ceftazidime, colistin and metronidazole.

The G+C content of the genome is 50.39%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers KC193239 and CBQR000000000, respectively. The type strain G5^T (= CSUR P290 = DSM 27179) was isolated from the fecal flora of a *Gorilla gorilla gorilla* from Cameroon.

Acknowledgements

Fadi Bittar was supported by a Chair of Excellence IRD provided by the Institut de Recherche pour le Développement/Méditerranée-Infection foundation. Mamadou Bhoye Keita was funded by the Méditerranée-Infection Foundation. The authors thank the Xegen company for automating the genome annotation process.

References

1. Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, Bittar F, Fournous G, Gimenez G, Maraninchi M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 2012; **18**:1185-1193. [PubMed](#)
2. Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, Raoult D. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *Eur J Clin Microbiol Infect Dis* 2013; **32**:637-645. [PubMed](#) <http://dx.doi.org/10.1007/s10096-012-1787-3>
3. Pfleiderer A, Lagier JC, Armougom F, Robert C, Viallettes B, Raoult D. Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. *Eur J Clin Microbiol Infect Dis* 2013; **32**:1471-1481. [PubMed](#) <http://dx.doi.org/10.1007/s10096-013-1900-2>
4. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Clostridium senegalense* sp. nov. *Stand Genomic Sci* 2012; **6**:386-395. [PubMed](#)
5. Lagier JC, Armougom F, Mishra AK, Ngyuen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. *Stand Genomic Sci* 2012;

- 6:315-324. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.2685971>
6. Lagier JC, El Karkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. *Stand Genomic Sci* 2012; **6**:116-125. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.2415480>
7. Roux V, El Karkouri K, Lagier JC, Robert C, Raoult D. Non-contiguous finished genome sequence and description of *Kurthia massiliensis* sp. nov. *Stand Genomic Sci* 2012; **7**:221-232. [PubMed](#) <http://dx.doi.org/10.4056/sigs.3206554>
8. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. *Stand Genomic Sci* 2012; **7**:1-11. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.2956294>
9. Hugon P, Ramasamy D, Lagier JC, Rivet R, Couderc C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Alistipes obesi* sp. nov. *Stand Genomic Sci* 2013; **7**:427-439. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3336746>
10. Ramasamy D, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Dielma fastidiosa* gen. nov., sp. nov., a new member of the Family *Erysipelotrichaceae*. *Stand Genomic Sci* 2013; **8**:336-351. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3567059>
11. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Genome sequence and description of *Timonella senegalensis* gen. nov., sp. nov., a new member of the suborder *Micrococcinae*. *Stand Genomic Sci* 2013; **8**:318-335. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3476977>
12. Mishra AK, Hugon P, Lagier JC, Nguyen TT, Couderc C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Enorma massiliensis* gen. nov., sp. nov., a new member of the Family *Coriobacteriaceae*. *Stand Genomic Sci* 2013; **8**:290-305. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3426906>
13. Ramasamy D, Lagier JC, Gorlas A, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus massiliogenes* sp. nov. *Stand Genomic Sci* 2013; **8**:264-278. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3496989>
14. Hugon P, Mishra AK, Lagier JC, Nguyen TT, Couderc C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Brevibacillus massiliensis* sp. nov.
- Stand Genomic Sci 2013; **8**:1-14. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.3466975>
15. Keita MB, Diene S, Robert C, Raoult D, Fournier PE, Bittar F. Non-contiguous finished genome sequence and description of *Bacillus massiliogorillae* sp. nov. *Stand Genomic Sci* 2013; **9**:93-105. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.4388124>
16. De Vos P, Ludwig W, Schleifer KH, Whitman WB. Family IV. *Paenibacillaceae* fam. nov. In Bergey's Manual of Systematic Bacteriology, 2nd Edition, vol 3 (The Firmicutes), Springer; New York, 2009, p. 269.
17. Abstract for the family *Paenibacillaceae*. NamesforLife, LLC. Retrieved October 21, 2013.
<http://doi.namesforlife.com/10.1601/tx.5108>
18. Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. *Antonie van Leeuwenhoek* 1993; **64**:253-260. [PubMed](#)
<http://dx.doi.org/10.1007/BF00873085>
19. The type species of the genus *Paenibacillus* Ash et al. 1994 is *Paenibacillus polymyxa*. Opinion 77. Judicial Commission of the International Committee on Systematics of Prokaryotes. *Int J Syst Evol Microbiol* 2005; **55**:513.
20. Zaitsev GM, Tsitko IV, Rainey FA, Trotsenko YA, Uotila JS, Stackebrandt E, Salkinoja-Salonen MS. New aerobic ammonium-dependent obligately oxalotrophic bacteria: description of *Ammoniphilus oxalaticus* gen. nov., sp. nov. and *Ammoniphilus oxalivorans* gen. nov., sp. nov. *Int J Syst Bacteriol* 1998; **48**:151-163. [PubMed](#) <http://dx.doi.org/10.1099/00207713-48-1-151>
21. Shida O, Takagi H, Kadokawa K, Komagata K. Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. *Int J Syst Bacteriol* 1996; **46**:939-946. [PubMed](#)
<http://dx.doi.org/10.1099/00207713-46-4-939>
22. Touzel JP, O'Donohue M, Debeire P, Samain E, Breton C. *Thermobacillus xylanilyticus* gen. nov., sp. nov., a new aerobic thermophilic xylan-degrading bacterium isolated from farm soil. *Int J Syst Evol Microbiol* 2000; **50**:315-320. [PubMed](#) <http://dx.doi.org/10.1099/00207713-50-1-315>
23. Saha P, Krishnamurthi S, Bhattacharya A, Sharma R, Chakrabarti T. *Fontibacillus aquaticus* gen. nov., sp. nov., isolated from a warm spring. *Int J Syst Evol Microbiol* 2010; **60**:422-428. [PubMed](#)
<http://dx.doi.org/10.1099/ijss.0.012633-0>

24. Kämpfer P, Rosselló-Mora R, Falsen E, Busse HJ, Tindall BJ. *Cohnella thermotolerans* gen. nov., sp. nov., and classification of 'Paenibacillus hongkongensis' as *Cohnella hongkongensis* sp. nov. *Int J Syst Evol Microbiol* 2006; **56**:781-786. [PubMed](http://dx.doi.org/10.1099/ijss.0.63985-0) <http://dx.doi.org/10.1099/ijss.0.63985-0>
25. Rivas R, García-Fraile P, Zurdo-Piñeiro JL, Mateos PF, Martínez-Molina E, Bedmar EJ, Sánchez-Raya J, Velázquez E. *Saccharibacillus sacchari* gen. nov., sp. nov., isolated from sugar cane. *Int J Syst Evol Microbiol* 2008; **58**:1850-1854. [PubMed](http://dx.doi.org/10.1099/ijss.0.65499-0) <http://dx.doi.org/10.1099/ijss.0.65499-0>
26. Collins MD, Lawson PA, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, Cai J, Hippe H, Farrow JA. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol* 1994; **44**:812-826. [PubMed](http://dx.doi.org/10.1099/00207713-44-4-812) <http://dx.doi.org/10.1099/00207713-44-4-812>
27. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archae, Bacteria, and Eukarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](http://dx.doi.org/10.1073/pnas.87.12.4576) <http://dx.doi.org/10.1073/pnas.87.12.4576>
28. Gibbons NE, Murray RGE. Proposals concerning the higher taxa of *Bacteria*. *Int J Syst Bacteriol* 1978; **28**:1-6. <http://dx.doi.org/10.1099/00207713-28-1-1>
29. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.
30. Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG (ed), Bergey's Manual of Systematic Bacteriology, First Edition, Volume 1, The Williams and Wilkins Co., Baltimore, 1984, p. 31-34.
31. Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 19-20.
32. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; **60**:469-472. <http://dx.doi.org/10.1099/ijss.0.022855-0>
33. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. <http://dx.doi.org/10.1099/00207713-30-1-225>
34. Prévôt AR. In: Hauderoy P, Ehringer G, Guillot G, Magrou J., Prévôt AR, Rosset D, Urbain A (eds), Dictionnaire des Bactéries Pathogènes, Second Edition, Masson et Cie, Paris, 1953, p. 1-692
35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](http://dx.doi.org/10.1038/75556) <http://dx.doi.org/10.1038/75556>
36. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011; **28**:2731-2739. [PubMed](http://dx.doi.org/10.1093/molbev/msr121) <http://dx.doi.org/10.1093/molbev/msr121>
37. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; **33**:152-155.
38. Bosshard PP, Zbinden R, Altweig M. *Paenibacillus turicensis* sp. nov., a novel bacterium harbouring heterogeneities between 16S rRNA genes. *Int J Syst Evol Microbiol* 2002; **52**:2241-2249. [PubMed](http://dx.doi.org/10.1099/ijss.0.02105-0) <http://dx.doi.org/10.1099/ijss.0.02105-0>
39. Kim DS, Bae CY, Jeon JJ, Chun SJ, Oh HW, Hong SG, Baek KS, Moon EY, Bae KS. *Paenibacillus elgii* sp. nov., with broad antimicrobial activity. *Int J Syst Evol Microbiol* 2004; **54**:2031-2035. [PubMed](http://dx.doi.org/10.1099/ijss.0.02414-0) <http://dx.doi.org/10.1099/ijss.0.02414-0>
40. Ueda J, Yamamoto S, Kurosawa N. *Paenibacillus thermoaerophilus* sp. nov., a moderately thermophilic bacterium isolated from compost. *Int J Syst Evol Microbiol* 2013; **63**:3330-3335. [PubMed](http://dx.doi.org/10.1099/ijss.0.048090-0) <http://dx.doi.org/10.1099/ijss.0.048090-0>
41. Lee FL, Kuo HP, Tai CJ, Yokota A, Lo CC. *Paenibacillus taiwanensis* sp. nov., isolated from soil in Taiwan. *Int J Syst Evol Microbiol* 2007; **57**:1351-1354. [PubMed](http://dx.doi.org/10.1099/ijss.0.64764-0) <http://dx.doi.org/10.1099/ijss.0.64764-0>
42. Takebe F, Hirota K, Nodasaka Y, Yumoto I. *Brevibacillus nitrificans* sp. nov., a nitrifying bacterium isolated from a microbiological agent for enhancing microbial digestion in sewage treatment tanks. *Int J Syst Evol Microbiol* 2012; **62**:2121-2126. [PubMed](http://dx.doi.org/10.1099/ijss.0.032342-0) <http://dx.doi.org/10.1099/ijss.0.032342-0>
43. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](http://dx.doi.org/10.1038/nbt1360) <http://dx.doi.org/10.1038/nbt1360>
44. Prodigal. <http://prodigal.ornl.gov>

45. GenBank database.
<http://www.ncbi.nlm.nih.gov/genbank>
46. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA gene in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. [PubMed](#) <http://dx.doi.org/10.1093/nar/25.5.0955>
47. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm160>
48. Lechner M, Findeis S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011; **12**:124. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-12-124>